**基于嵌入系统与深度学习的**

**人体姿态和生理状态视频监测系统**

**开题报告**

专业综合实践训练2021

生医8 李浩伟 李子涵 江柔蓝

# 一、项目背景

人口老龄化是世界性问题，在中国的发展尤为迅速。国家统计局发布的《中华人民共和国2019年国民经济和社会发展统计公报》显示，中国60周岁及以上的老年人口数量已达到2.5亿，占人口总数的18.1%，其中65周岁以上占人口总数12.6%[1]，这些数据说明中国已经进入老龄化社会。社会各界都十分重视养老问题，给健康监护及远程医疗领域带来了更高的机遇和挑战。

家庭养老是当前养老模式的主流之一。在中国特有的“4+2+1”家庭结构下，子女多因工作无暇照顾老人，老人独居在家中通常面临着一系列的问题，一方面其在生活中更容易出现突发的危险，如摔倒等，而这些事件的发生往往会对老龄人的身体健康产生很严重的影响，或者预示着老年人的健康状况已经出现了重要变化；另一方面，老年人的日常作息，如一天中坐下、站立、躺卧、行走时间会对老年人的身体健康产生很重要的影响，也能反映老年人的健康状况，这些统计数据将为“大健康”战略的发展提供支撑。

因此，我们希望设计开发一个基于视频的多功能健康状态监测系统，重点实现对老年人的行为模式的实时识别、统计和反馈，为家庭养老的安全和健康提供助力。

近年来，家用监控设备已经被相当多的家庭所接受，全球著名消费科技市场研究机构Strategy Analytics发布最新的研究报告《2019年智能家居监控摄像头市场预测和分析》指出，2019年全球消费者将在智能家居摄像头上支出近80亿美元；在美国、英国、德国和法国中，有将近三分之一的家庭安装了监控摄像头。在中国，选择安装家用监控摄像头的家庭也已经超过10%，并且正在迅速增长。由此我们认为，基于视频的健康检测系统是能够被大部分家庭所接受的。

为了实现对人体姿态的检测识别，目前较为常见的系统可以分为基于视频图像和基于可穿戴式传感器两类[2][3]。与基于可穿戴式传感器的姿态检测系统相比，基于视频的方法具有“非接触”的优点，安装方便，能够提供长时间、稳定的检测，有效避免传感器丢失、损坏、电量不足等情况，且不会对老人的日常生活造成不便。另一方面，基于视频的方法可以一次性采集多种参数，相当于在被观测者的不同部位同时佩戴了多组传感设备，且具有很强的可拓展性，可以迅速将计算机视觉领域的新兴研究应用在该系统中。此外，在“居家养老”这一特定的应用场景下，室内环境和周围光线的变化较为简单，非常适合视频检测系统的使用。

# 二、技术背景

项目的核心技术是人体行为模式的提取。对于这个目标现存许多不同的解决方案。包括基于各种传感器的检测和基于视频的检测等等。视频检测因为不需要传感器佩戴，方便子女进行实时的照看等等特性具有较好的探索潜力。普遍得讲，行为模式得识别一般分为人体追踪和模式识别两个部分。在前一部分即存在光流法等纯视频分析方法，也有open pose，yolo，RCNN等深度学习方法。而模式识别较为广泛的方式为先进行骨架或特征点的提取，而后通过特征点的空间信息进行识别。NTU-RGBD数据集，作为一个广泛使用的行为识别的数据集，在近些年已经验证了基于骨架图的模式识别在正确率上可以达到一个非常高的水平，这些网络通常都是基于图网络的结构。但是这些都是对于本地视频的行为检测，对于实时的检测，问题则困难的多。近些年有人提出一些轻量级网络，实现了高达180fps的检测，虽然这些网络往往以一些准确度的牺牲为代价，但是其预示着附带端处理功能的小型化的家用行为检测视频设备是可能的。

地平线作为一家国产公司，在计算机视觉方面有着出色的成就，和众多的智能汽车厂商保持着紧密的合作，提供智能驾驶的解决方案。其提供的开发板X3在2.5W的功耗下达到了2TFLOPs/W的高计算能耗比（现有设备中仅有haswell-eq 2.13，Raspberry 2.02）比其略微优秀一些。但是X3在视频编解码速度上占有一定的优势，同时其作为一个放释放的产品，也相应具有更大的发展潜能。因此我们选择该设备作为我们的硬件基础。

脉搏波信号是现在非常常用的获取心率数据的手段。投射式和反射式的光电容积脉搏波测量方法被广泛应用在接触式的心率检测手段中。这种光电容积法本质上是通过主动光的方式获得血管容积变化的信息。但是在现实中，我们并不需要主动光，而在环境中天然得存在我们需要的颜色的光线。因此我们可以通过捕捉人脸对一定波长光线的反射信号捕捉心率。这个过程中诞生了非常多的方法，包括绿色光滤波，HSV空间检测，独立成分分析等等。但是我们可以发现现有的大量论文都是基于静态、稳定、补光的条件下进行的。实际上，人脸上的反射光的变化主要由三个变化导致：环境光的变化，人脸对环境光的反射变化和脉搏波引导的光吸收变化。所以有研究者希望对人脸的三个不同的独立部位进行捕捉，通过独立成分分析的方式对脉搏波信号进行分离。同时，在现有研究中，为了捕捉频率尽量高，一般对于人脸不同部分的识别仅采用haar cascade的方式，其在一些时候会产生较大的偏差。而这个问题恰恰好在上面一部分得到了解决：X3的开发板提供了非常好的算力支撑可以告诉获取部位信息，同时可以告诉对视频信号进行处理。我们希望在这个基础上使用，并尽量优化前人已有的检测算法。同时此处，DEAP数据集提供了大量的视频和心电数据，我们正在争取获得该数据集，以对算法的结果进行更好的验证。

核心硬件

地平线X3开发者套件（由王广志老师实验室提供）

* 四核ARM [Cortex-A53@1.5GHz](mailto:Cortex-A53@1.5GHz)
* [双核BPU@1.2GHz](mailto:双核BPU@1.2GHz) 5TFLOPS等效算力vs 3090 35.7TFLOPS
* 2GB 三星 LPDDR4 + 8G eMMC5.1
* Hdmi, TF/SD, MIPI CSI, MIPI DSI, UART, 40 pin接口

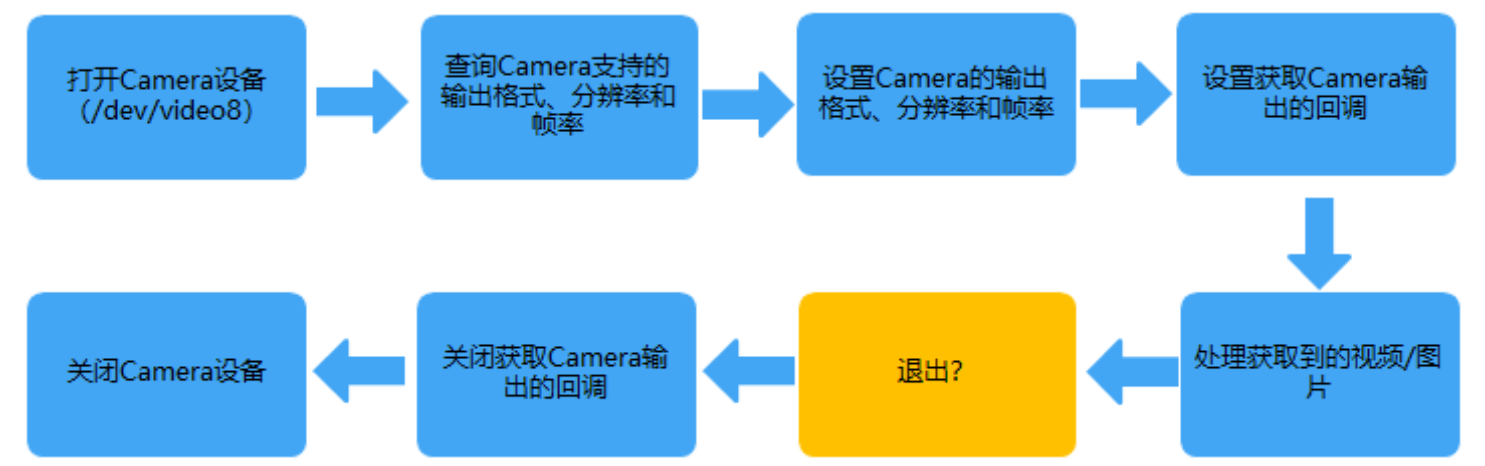
硬件/软件需求

* Linux开发机或虚拟机，Ubuntu或Centos，用来进行AIExpress编译和调试开发
* Windows PC，调试串口，烧录程序和系统
* X3sdb开发板
* 显卡服务器，用以进行模型训练
* 灰点高速相机(1.3MP, 149fps)

开发流程

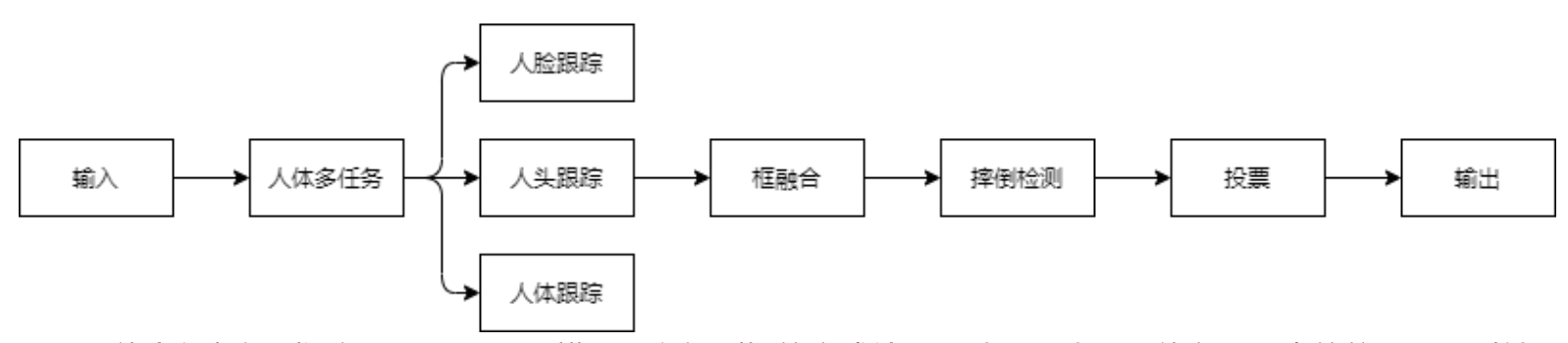
1. 开发板系统烧录，进行基本的系统配置
2. Linux中编写程序，进行调试、利用提供的开发板虚拟软件进行模拟运行。
3. 利用神经网络训练模型，并进行测试（与1同步）
4. 编写基于人脸视频的心率检测代码，并进行测试（与1同步）
5. 软件编译
6. 通过烧录软件或者scp协议把编译好的结果部署在开发板上
7. 运行，并通过网口将数据传输到web端进行展示

一般基于Camera的软件执行流程



(图源地平线AI论坛, zhuk, <https://developer.horizon.ai/forum/id=5f312d96cc8b1e59c8581511)>

已有的摔倒检测的基本架构(fast RCNN+CNN)



(图源地平线AI论坛, asdf, https://developer.horizon.ai/forum/id=5efab48f38ca27ba028078dd)

在本项目中该流程需要处理的事情（后几个在有时间的情况下完成）

1. 获得每一帧图片的时间信息，帧内的骨架信息，帧内识别人脸部分对应的rgb的图像
2. 实现更多的基本分类，对视频中的多个骨架实现行为模式识别
3. 通过面部图像进行颜色特征提取，通过一些基本的生理学假设、建模和滤波进行PPG的分离，实现心率检测
4. 基于视频进行人脸识别，达成行为模式统计的作用
5. 基于上述的统计结果设计一定的交互
6. 通过多个摄像头的骨架进行三维重建，进行更精确的复原

项目难点

* 开发板推出时间晚(2020/9/9)，开发社区建构不完善，开发工具集少，开发手段较为局限，在安装调试的时候没有好的的经验或者案例可以借鉴，需要更多得自己探索或和官方开发人员进行沟通
* 开发板现有得硬件支持技术并不成熟，相比而言，其提供了较高得算力，为轻量级的模式识别提供了可能性，但是另一方面其支持的数据格式，硬件种类都 较少。
* 需要交叉编译，多平台联合开发

项目时间安排

第八周前：文献调研，软件系统准备，服务器基本配置

9-11周：开发板接口调试，视频心率检测，行为模式识别算法构架搭建

12-14周：神经网络训练，算法测试，软硬件联调，GUI编写

15-16周：总结，完成前述中潜在的遗漏部分，尽可能完成上述提到的提高部分，报告和PPT撰写

# 参考文献

[1]国家统计局.中华人民共和国2019年国民经济和社会发展统计公报[N]. 2020-2-28.

[2]胡双杰. 基于惯性传感器的实时跌倒检测算法研究与系统设计[D].上海交通大学,2019.

[3]吴聪. 基于惯性传感器的老年人姿态监测系统设计[D].山东大学,2019.